

**IMPLEMENTASI ALGORITMA *RABIN-KARP* DAN *STEMMING*
NAZIEF ADRIANI UNTUK PENDETEKSI PLAGIARISME
ABSTRAKSI SKRIPSI**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh
Andy Saputra
15.11.9057

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2018**



PERSETUJUAN

SKRIPSI

IMPLEMENTASI ALGORITMA *RABIN-KARP* DAN *STEMMING*

NAZIEF ADRIANI UNTUK PENDETEKSI PLAGIARISME

ABSTRAKSI SKRIPSI

yang dipersiapkan dan disusun oleh

Andy Saputra

15.11.9057

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 Mei 2018

Dosen Pembimbing,



Anggit Dwi Hartanto, M.Kom.

NIK. 190302163

PENGESAHAN

SKRIPSI

IMPLEMENTASI ALGORITMA *RABIN-KARP* DAN *STEMMING*

NAZIEF ADRIANI UNTUK PENDETEKSI PLAGIARISME

ABSTRAKSI SKRIPSI

yang dipersiapkan dan disusun oleh

Andy Saputra

15.11.9057

telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Agustus 2018

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Agus Fatkhurohman, M.Kom

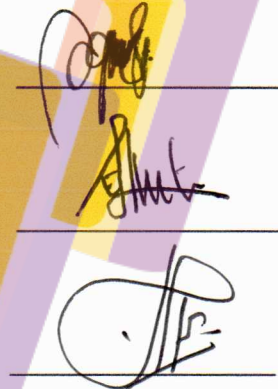
NIK. 190302249

Erni Seniwati, S.Kom, M.Cs

NIK. 190302231

Anggit Dwi Hartanto, M.Kom

NIK. 190302163



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
tanggal 20 Agustus 2018

DEKAN FAKULTAS ILMU KOMPUTER



Krisnawati, S.Si, MT

NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 20 Agustus 2018



Andy Saputra

NIM. 15.11.9057

MOTTO

Boleh jadi kamu membenci sesuatu, padahal ia amat baik bagi kamu. Dan boleh jadi kamu mencintai sesuatu, padahal ia amat buruk bagi kamu. Allah Maha mengetahui sedangkan kamu tidak mengetahui” (Al-Baqarah: 216)

Untuk Setiap 1 menit yang kamu habiskan untuk kemarahan , kamu kehilangan 60 detik kebahagiaan



PERSEMBAHAN

Saya mempersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung maupun tidak langsung dalam proses pembuatan skripsi.

Tuhan Yang Maha Esa yang memberikan segala nikmat dan kasih sayangnya sampai sejauh ini.

1. Kedua orang tua saya, yang selalu mendoakan, membuat semangat dan memberikan semua fasilitas untuk penunjang kuliah.
2. Bapak Anggit Dwi Hartanto, M.kom. yang telah membimbing saya dari awal sampai akhir proses pembuatan skripsi.
3. Dosen-dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu selama kuliah.
4. Teman-Teman FAKE GRUP yang telah menemani dari awal kuliah sampai selesai. Semoga kita semua selalu bahagia dan menjadi pribadi yang lebih baik lagi.
5. Teman coding saya Akhid Bunayari dan Ardi Pratama yang telah mendoakan, ngoding bareng dan memberi semangat.
6. Dicky Suryo Saputro yang telah memberikan masukan dan saran dalam proses pembuatan skripsi.

KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat, hidayah dan kekuatan sehingga saya dapat menyelesaikan skripsi yang berjudul Implementasi Algoritma *Rabin-Karp* Dan *Stemming* Nazief Adriani Untuk Pendeteksi Plagiarisme Abstraksi Skripsi.

Skripsi ini saya buat guna menyelesaikan studi jenjang Strata Satu (S1) pada program studi Teknik Informatika fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Selain itu juga merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang program strata 1 dan untuk memperoleh gelar Sarjana Komputer.

Dengan selesainya skripsi ini, Maka pada kesempatan ini saya mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Krisnawati, S.Si, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Bapak Anggit Dwi Hartanto, M.kom. selaku dosen pembimbing yang selalu bijaksana memberikan bimbingan, nasehat serta waktunya selama penulisan skripsi ini.
4. Dosen Penguji (Agus Fatkhurohman, M.Kom. Emi Seniwati, S.Kom, M.Cs) dan segenap Dosen dan Karyawan Universitas AMIKOM Yogyakarta yang telah berbagi ilmu dan pengalamannya.
5. Kedua orang tua saya yang telah mendoakan, mendukung dan memberikan semangat.
6. Saudara-saudara yang senantiasa mendoakan, memberi semangat dan bantuan.
7. Teman-Teman FAKE GRUP yang telah menemani dari awal kuliah sampai selesai. Semoga kita semua selalu bahagia dan menjadi pribadi yang lebih baik lagi.
8. Teman coding saya Akhid Bunayari dan Ardi Pratama yang telah mendoakan, ngoding bareng dan memberi semangat.

9. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung.

Semoga Allah Subhanahu wata'ala memberikan balasan yang lebih kepada semua yang telah ikut membantu saya dan menyelesaikan skripsi ini. Demi perbaikan selanjutnya, saran dan kritik yang membangun akan diterima dengan senang hati dan rasa terima kasih. Semoga skripsi ini dapat bermanfaat bagi saya dan kita semua.

Yogyakarta, 20 Agustus 2018

Andy Saputra

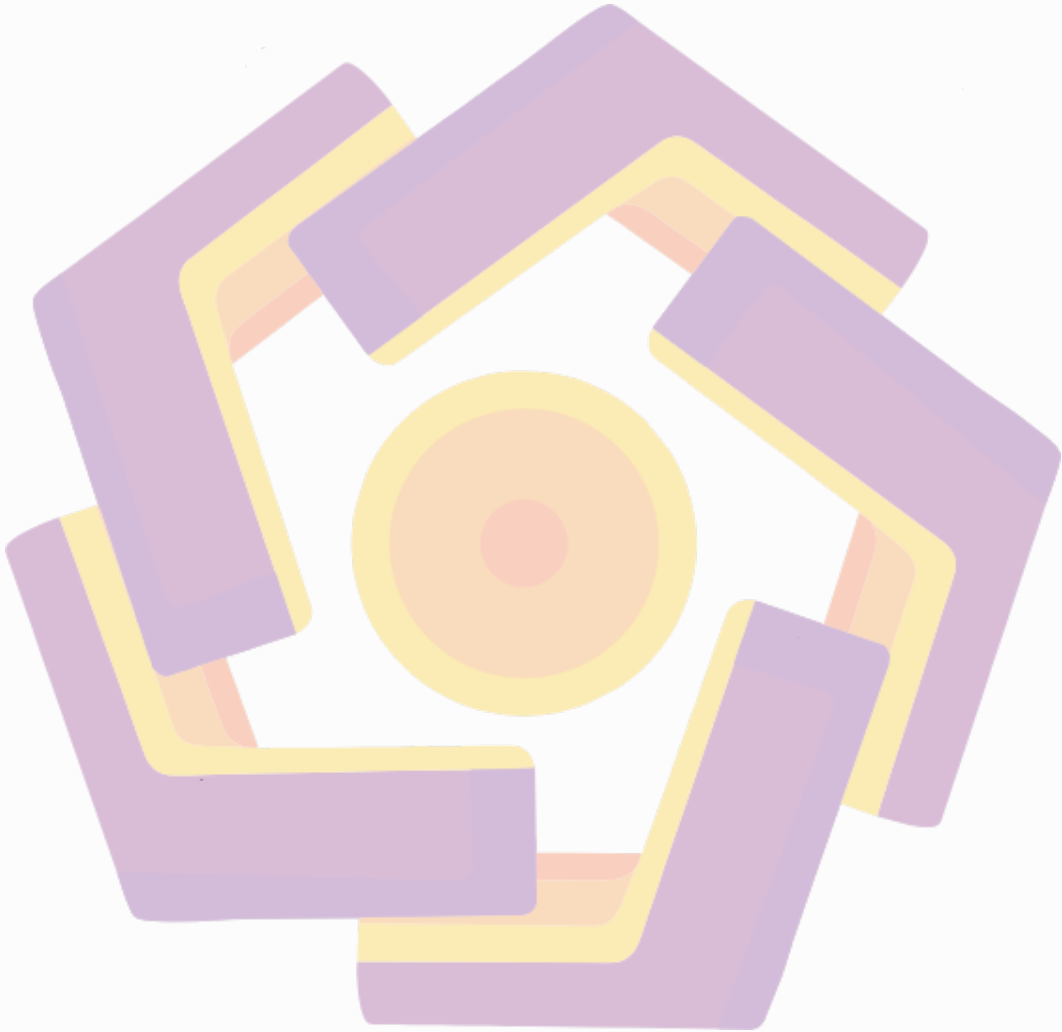
DAFTAR ISI

LEMBAR JUDUL.....	i
PERSETUJUAN.....	ii
PENGESAHAN.....	iii
PERNYATAAN.....	IV
MOTTO.....	V
PERSEMBAHAN.....	VI
KATA PENGANTAR.....	VII
DAFTAR ISI.....	IX
DAFTAR TABEL.....	XIII
DAFTAR GAMBAR.....	XIV
INTISARI.....	XVII
<i>ABSTRACT</i>	XVIII
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Maksud dan Tujuan Penelitian.....	4
1.4.1 Maksud.....	4
1.4.2 Tujuan.....	4
1.5 Manfaat Penelitian.....	5
1.6 Metode Penelitian.....	5
1.7 Sistematika Penulisan.....	7
BAB II LANDASAN TEORI.....	9

2.1	Kajian Pustaka.....	9
2.2	Dasar Teori.....	16
2.2.1	<i>Pengertian Plagiarisme</i>	16
2.2.1.1	Tipe Plagiarisme.....	17
2.2.1.2	Metode Pendeteksi Plarigarisme	17
2.2.2	<i>Text Mining</i>	19
2.2.2.1.	Tahapaan <i>Text Mining</i>	20
2.2.2.1.1	<i>Text Preprocessing</i>	20
2.2.2.1.2	<i>Text Transformation</i>	24
2.2.2.1.3	<i>Feature Selection</i>	24
2.2.2.1.4	<i>Pattern Discovery</i>	25
2.2.3	Algoritma <i>Stemming Nazief-Adriani</i>	25
2.2.4	<i>K-Gram</i>	29
2.2.5	<i>Hashing</i>	30
2.2.6	Algoritma <i>Rabin-Karp</i>	30
2.2.7	Pengukuran Nilai <i>Similarity</i>	33
2.2.8	<i>Object Oriented Analysis and Design Lifecycle</i>	34
2.2.9	Perancangan <i>UML</i>	34
2.2.10	<i>PHP</i>	40
2.2.11	<i>MySQL</i>	41
2.2.12	<i>JavaScript</i>	41
2.2.13	<i>Xampp</i>	41
BAB III ANALISIS DAN PERANCANGAN PENELITIAN		43
3.1	Identifikasi Masalah	43
3.2	Analisis Masalah	43
3.3	Hasil Analisis	44
3.4	Deskripsi Sistem.....	44
3.5	Perhitungan Manual <i>Rabin-Karp</i>	46
3.6	Analisis Kebutuhan	51
3.6.1	Analisis Kebutuhan <i>Fungsional</i>	51

3.6.1.1	Definisi Fitur	51
3.6.1.2	Aktor	53
3.6.1.3	<i>Use Case</i>	53
3.6.1.4	Skenario <i>Use Case</i>	54
3.6.2	<i>Analisis Kebutuhan Non-Fungsional</i>	72
3.6.2.1	Kebutuhan Perangkat Keras	72
3.6.2.2	Kebutuhan Perangkat Lunak	73
3.7	Perancangan Perangkat Lunak.....	73
3.7.1	<i>Activity Diagram</i>	73
3.7.2	<i>Sequence Diagram Aplikasi</i>	79
3.7.3	Rancangan Antarmuka Pengguna (<i>User Interface</i>).....	82
BAB IV	HASIL DAN PEMBAHASAN	87
4.1	Database dan Tabel	87
4.2	Implementasi Antarmuka Pengguna	89
4.3	Pembahasan Source Code.....	95
4.4	Pengujian Sistem.....	102
4.4.1	<i>White Box Testing</i>	102
4.4.2	<i>Black Box Testing</i>	103
4.5	Hasil Dan Analisis Pengujian	104
4.5.1	Uji Coba Akurasi Dan Waktu <i>Stemming</i> Nazief Adriani, Sastrawi, Tala, arifin setiono	105
4.5.2	Uji Coba Nilai <i>K-Gram</i> 4,5,6 dan Nilai Basis Prima 5,11,13.....	106
4.5.3	Uji Coba Perbandingan Dengan <i>Stemming</i> Sastrawi dan Nazief Adriani.....	109
4.5.4	Uji Coba Perbandingan Dengan <i>Stemming</i> Sastrawi Dan <i>Stemming</i> Nazief Adriani Dengan Mengubah Tata Bahasa	114
4.5.5	Uji Coba Waktu Dengan Nilai <i>K-Gram</i> 4,5,6 Dan Nilai Basis Prima 5,11,23.....	118
4.5.6	Perhitungan Dalam Sistem.....	119
BAB V	PENUTUP	122

5.1. Kesimpulan	122
5.2. Saran	124
DAFTAR PUSTAKA	125
LAMPIRAN	127



DAFTAR TABEL

Tabel 2.1 Perbedaan Penelitian	13
Tabel 2.2: <i>Use Case</i> Diagram.....	35
Tabel 2.3: <i>Sequence</i> Diagram.....	37
Tabel 2.4: <i>Class</i> Diagram.....	39
Tabel 3.1: Tabel Dokumen.....	46
Tabel 3.2: Tabel <i>Text Preprocessing</i>	47
Tabel 3.3: Tabel Hasil <i>Text Preprocessing</i>	47
Tabel 3.4 : Tabel <i>Parshing K-Gram</i> dengan nilai $k=4$	48
Tabel 3.5 : Tabel <i>Rolling Hash</i>	49
Tabel 3.6: Tabel <i>Similarity Rolling Hash</i>	50
Tabel 3.7 : Tabel Daftar Fitur Aplikasi.....	51
Tabel 3.8 : Tabel Daftar Aktor	53
Tabel 3.9 : Tabel Skenario <i>Use Case</i> Membandingkan Abstraksi Skripsi	55
Tabel 3.10 : Tabel Skenario <i>Use Case Text Preprocessing</i>	58
Tabel 3.11 : Tabel Skenario <i>Use Case</i> Algoritma <i>Rabin-Karp</i>	59
Tabel 3.12 : Tabel Skenario <i>Use Case</i> Hasil Cek Plagiarisme	61
Tabel 3.13 : Tabel Skenario <i>Use Case Login</i>	61
Tabel 3.14 : Tabel Skenario <i>Use Case</i> Pengaturan	62
Tabel 3.15 : Tabel Skenario <i>Use Case</i> Tambah Abstraksi	63
Tabel 3.16 : Tabel Skenario <i>Use Case</i> Ubah Abstraksi.....	65
Tabel 3.17 : Tabel Skenario <i>Use Case</i> Hapus Abstraksi	68
Tabel 3.18 : Tabel Skenario <i>Use Case</i> Data Abstraksi Skripsi	69
Tabel 3.19 : Tabel Skenario <i>Use Case</i> Data Abstraksi Uji.....	70
Tabel 3.20 : Tabel Skenario <i>Use Case</i> Ubah Admin.....	70
Tabel 3.21 : Tabel Skenario <i>Use Case</i> Data Admin.....	71
Tabel 3.22 : Tabel Skenario <i>Use Case</i> Data Admin.....	72
Tabel 4.1 : Tabel Hasil <i>Black Box</i> Testing.....	103
Tabel 4.2 : Tabel Hasil Uji <i>Stemming</i> Nazief Adriani, Sastrawi, Tala, Arifin Setiono.....	105

DAFTAR GAMBAR

Gambar 2.1 : Contoh proses <i>case folding</i>	21
Gambar 2.2 : Contoh proses <i>number removal</i>	21
Gambar 2.3 : Contoh proses <i>tokenizing</i>	22
Gambar 2.4 : Contoh proses <i>filtering</i>	22
Gambar 2.5 : Contoh proses <i>Stemming</i>	24
Gambar 2.5: <i>Flowchart</i> Algoritma <i>Stemming</i> Nazief-Adriani.....	27
Gambar 2.6 : Rumus <i>Rolling Hash</i>	30
Gambar 2.8 : Rumus <i>Rabin-Karp</i>	31
Gambar 2.10: <i>Fingerprint</i> awal[2]	32
Gambar 2.11: <i>Menggeser Fingerprint</i> [2].....	32
Gambar 2.12: Perbandingan kedua [2].....	32
Gambar 2.13: Perbandingan keempat(nilai <i>hash</i> sama) [2].....	33
Gambar 2.14: Perbandingan keempat(nilai <i>hash</i> sama) [2].....	33
Gambar 2.15: Rumus Dice's Similarity Coefficient.....	34
Gambar 3.1 : Gambar Diagram Alir Proses Pendeteksi Plagiarisme	45
Gambar 3.2 Gambar <i>Use Case Diagram</i> Aplikasi.....	54
Gambar 3.3 : Gambar <i>Activity Diagram</i> Membandingkan Abstraksi Skripsi	74
Gambar 3.4 : Gambar <i>Activity Diagram</i> Text Preprocessing	74
Gambar 3.5 : Gambar <i>Activity Diagram</i> <i>Rabin-Karp</i>	75
Gambar 3.6 : Gambar <i>Activity Diagram</i> <i>Login</i>	75
Gambar 3.7 : Gambar <i>Activity Diagram</i> Pengaturan	76
Gambar 3.8 : Gambar <i>Activity Diagram</i> Ubah Admin.....	76
Gambar 3.9 : Gambar <i>Activity Diagram</i> Tampil Cek Plarigarisme.....	77
Gambar 3.10 : Gambar <i>Activity Diagram</i> Tambah Abstraksi Skripsi	77
Gambar 3.11 : Gambar <i>Activity Diagram</i> Ubah Abstraksi Skripsi.....	78
Gambar 3.12 : Gambar <i>Activity Diagram</i> Hapus Abstraksi Skripsi	78
Gambar 3.13 : Gambar <i>Sequence Diagram</i> Membandingkan Abstraksi Skripsi ..	79
Gambar 3.14 : Gambar <i>Sequence Diagram</i> <i>Login</i> Admin	80
Gambar 3.15 : Gambar <i>Sequence Diagram</i> Ubah Admin.....	80

Gambar 3.16 : Gambar <i>Sequence</i> Diagram Pengaturan.....	81
Gambar 3.17 : Gambar <i>Sequence</i> Diagram Tampil Data Cek Uji Plagiarisme	81
Gambar 3.18 : Gambar <i>Sequence</i> Diagram Abstraksi (Tampil,Tambah,Ubah,Hapus Data).....	82
Gambar 3.21 : Gambar Form Cek Plagiarisme	83
Gambar 3.22 : Gambar Halaman Hasil Plagiarisme.....	83
Gambar 3.23 : Gambar Form <i>Login</i> Admin.....	84
Gambar 3.24 : Gambar Halaman Dashboard	84
Gambar 3.25 : Gambar Halaman Data Abstraksi.....	84
Gambar 3.26 : Gambar Form Tambah Data Abstraksi.....	85
Gambar 3.27 : Gambar Form Edit Data Abstraksi	85
Gambar 3.28 : Gambar Halaman Data Uji Plagiarisme.....	85
Gambar 3.29 : Gambar Halaman Admin	86
Gambar 3.30 : Gambar Form Edit Admin	86
Gambar 3.31 : Gambar Halaman Pengaturan.....	86
Gambar 4.1 : Gambar Implementasi tabel abstrak_asli	87
Gambar 4.2 : Gambar Implementasi tabel abstrak_uji	88
Gambar 4.3 : Gambar Implementasi tabel admin.....	88
Gambar 4.5 : Gambar Implementasi tabel temp_hasil	88
Gambar 4.6 : Gambar Implementasi tabel hasil_uji	89
Gambar 4.7 : Gambar Implementasi tabel pengaturan	89
Gambar 4.7 : Gambar Halaman form cek plagiarisme	90
Gambar 4.8 : Gambar hasil cek plagiarisme	90
Gambar 4.9 : Gambar Halaman form <i>Login</i>	91
Gambar 4.10 : Gambar Halaman <i>Dashboard</i>	91
Gambar 4.11 : Gambar Halaman abstrak.....	92
Gambar 4.12 : Gambar Halaman form tambah abstrak	92
Gambar 4.13 : Gambar Halaman form edit abstrak.....	93
Gambar 4.14 : Gambar Halaman hasil uji cek plagiarisme	93
Gambar 4.15 : Gambar Halaman admin	94
Gambar 4.15 : Gambar Halaman edit admin.....	94

Gambar 4.16 : Gambar Halaman pengaturan.....	95
Gambar 4.17 : Gambar Diagram dengan nilai <i>K-Gram</i> 4.....	106
Gambar 4.18 : Gambar Diagram dengan nilai <i>K-Gram</i> 5.....	107
Gambar 4.19 : Gambar Diagram dengan nilai <i>K-Gram</i> 6.....	107
Gambar 4.20 : Gambar Diagram dengan nilai <i>K-Gram</i> 4.....	110
Gambar 4.21 : Gambar Diagram perbandingan waktu proses <i>Stemming</i> sastrawi dan <i>Stemming</i> nazief adriani.....	110
Gambar 4.22 : Gambar Diagram dengan nilai <i>K-Gram</i> 5.....	111
Gambar 4.23 : Gambar Diagram perbandingan waktu proses <i>Stemming</i> sastrawi dan <i>Stemming</i> nazief adriani.....	111
Gambar 4.24 : Gambar Diagram dengan nilai <i>K-Gram</i> 6.....	112
Gambar 4.25 : Gambar Diagram perbandingan waktu proses <i>Stemming</i> sastrawi dan <i>Stemming</i> nazief adriani.....	112
Gambar 4.26 : Gambar Diagram dengan nilai <i>K-Gram</i> 4.....	114
Gambar 4.27 : Gambar Diagram perbandingan waktu proses <i>Stemming</i> sastrawi dan <i>Stemming</i> nazief adriani.....	115
Gambar 4.28 : Gambar Diagram dengan nilai <i>K-Gram</i> 5.....	115
Gambar 4.29 : Gambar Diagram perbandingan waktu proses <i>Stemming</i> sastrawi dan <i>Stemming</i> nazief adriani.....	116
Gambar 4.30 : Gambar Diagram dengan nilai <i>K-Gram</i> 6.....	116
Gambar 4.31 : Gambar Diagram perbandingan waktu proses <i>Stemming</i> sastrawi dan <i>Stemming</i> nazief adriani.....	117
Gambar 4.32 : Gambar Uji Coba Waktu.....	118
Gambar 4.33 : Gambar form Input <i>K-Gram</i> dan Teks	120
Gambar 4.34 : Gambar Hasil Case Folding Dan Filtering (Stopword Removal)	120
Gambar 4.35 : Gambar Hasil <i>Stemming</i> dan Hapus Spasi.....	120
Gambar 4.36 : Gambar Hasil <i>K-Gram</i> dan <i>Rolling Hash</i>	121
Gambar 4.37 : Gambar Hasil <i>Rolling Hash</i> Sama dan Hasil <i>Similarity</i>	121

INTISARI

Plagiarisme adalah penjiplakan atau pengambilan karangan, pendapat, dan sebagainya dari orang lain dan menjadikannya seolah karangan dan pendapat sendiri. Plagiat dapat dianggap sebagai tindak pidana karena mencuri hak cipta orang lain. Sudah di tulis dalam undang undang Pasal 2 ayat (1) berbunyi “Hak Cipta merupakan hak eksklusif bagi Pencipta atau Pemegang Hak Cipta untuk mengumumkan atau memperbanyak Ciptaannya, yang timbul secara otomatis setelah suatu ciptaan dilahirkan tanpa mengurangi pembatasan menurut peraturan perundang-undangan yang berlaku”.

Penggunaan Algoritma *Stemming* Nazief dan Adriani dalam mencari kata dasar untuk setiap kata dalam abstraksi skripsi yang di uji. Setelah didapat kata dasar untuk setiap kata, di bentuk gram dan di cari nilai hash menggunakan algoritma *Rabin-Karp*. Kemudian nilai hash abstraksi skripsi yang di uji akan di bandingkan dengan nilai hash abstraksi skripsi yang ada di database.

Hasil yang diharapkan dapat mengetahui seberapa tingkat plagiarisme abstraksi skripsi dengan mekanisme pengujian mencari persentase nilai *similarity* abstraksi skripsi yang di bandingkan dengan abstraksi skripsi dari database dan hasil perbandingannya dalam bentuk persen, disimpan di database dan ditampilkan di sistem secara berurut dari total persentase kemiripan dari yang terbesar sampai terkecil.

Kata Kunci: algoritma *rabin-karp*, *Stemming* nazief adriani, plagiarisme, kecerdasan buatan

ABSTRACT

Plagiarism is the plagiarism or the making of articles, opinions, etc. from others and makes it as an article and an opinion of its own. Plagiarism can be regarded as a crime for stealing the copyrights of others. Already written in the Undang Undang Pasal 2 Ayat (1) reads "*Hak Cipta merupakan hak eksklusif bagi Pencipta atau Pemegang Hak Cipta untuk mengumumkan atau memperbanyak Ciptaannya, yang timbul secara otomatis setelah suatu ciptaan dilahirkan tanpa mengurangi pembatasan menurut peraturan perundang-undangan yang berlaku*".

Use of *Stemming nazief adriani* in finding the basic word for every word in abstraction essay in the test. Having obtained the basic word for each word, formed gram and searched hash value using rabin karp algorithm. Then the hash value abstraction essay tested in comparison with the hash value abstraction essay from the database.

The expected result can know how level plagiarism abstraction essay with the testing mechanism to find the percentage value similarity abstraction essay that in comparison with abstraction essay from the database and the results of the comparison in the percent, stored in the database and displayed in the system sequentially from the total percentage of similarity form the largest to the smallest.

Keyword: rabin-karp algorithm, *nazief adriani* Stemming, plagiarism, artificial intelligence